Genetics of human cardiovascular traits

Verweij, Niek

6.

FIFTY-TWO GENETIC LOCI INFLUENCING HUMAN ELECTRICALLY ACTIVE MYOCARDIAL MASS

Pim van der Harst*, Jessica van Setten*, Niek Verweij*, Georg Vogler*, Lude Franke*, Matthew Maurano*, Xinchen Wang*, Irene Mateo Leach*, Mark Eijgelsheim, Nona Sotoodehnia, Caroline Hayward, Rossella Sorice, Osorio Meirelles, Leo-Pekka Lyytikäinen, Ozren Polašek, Toshiko Tanaka, Dan Arkin, Sheila Ulivi, Stella Trompet, Martina Müller-Nurasyid, Albert V. Smith, Marcus Dörr, Kathleen F. Kerr, Jared W. Magnani, Fabiola Del Greco M., Weihua Zhang, Ilja M. Nolte, Claudia T. Silva, Sandosh Padmanabhan, Vinicius Tragante, Tõnu Esko, Goncalo R. Abecasis, Karl Andersen, Phil Barnett, Josh C. Bis, Rolf Bodmer, Brendan M. Buckley, Harry Campbell, Megan V. Cannon, Aravinda Chakravarti, Lin Y. Chen, Alessandro Delitala, Richard B. Devereux, Pieter A. Doevendans, Francisco S. Domingues, Anna F. Dominiczak, Luigi Ferrucci, Ian Ford, Christian Gieger, Tamara B. Harris, Eric Haugen, Dena G. Hernandez, Hans L. Hillege, Albert Hofman, Annamaria Iorio, Mika Kähönen, Ivana Kolcic, Ishminder K. Kooner, Jaspal S. Kooner, Jan A. Kors, Edward G. Lakatta, Kasper Lage, Lenore J. Launer, Daniel Levy, Alicia Lundby, Peter MacFarlane, Dalit May, Thomas Meitinger, Andres Metspalu, Stefania Nappo, Silvia Naitza, Shane Neph, Alex S. Nord, Tersa Nutile, Peter M. Okin, Jesper V. Olsen, Ben A. Oostra, Gina Peloso, Josef M. Penninger, Len A. Pennacchio, Siegried Perz, Annette Peters, Arne Pfeufer, Maria Grazia Pilia, Peter P. Pramstaller, Bram P. Prins, Olli T. Raitakari, Soumya Raychaudhuri, Ken M. Rice, Jerome I. Rotter, David Schlessinger, Carsten O. Schmidt, Jobanpreet Sehmi, Herman H.W. Silljé, Gianfranco Sinagra, Moritz F. Sinner, Kamil Slowikowski, Elsayed Z. Soliman, Timothy D. Spector, Wilco Spiering, John A. Stamatoyannopoulos, Ronald P. Stolk, Konstantin Strauch, Sian-Tsung Tan, Kirill V. Tarasov, Bosco Trinh, Andre G. Uitterlinden, Malou van den Boogaard, Cornelia M. van Duijn, Wiek H. van Gilst, Jorma S. Viikari, Peter Visscher, Veronique Vitart, Uwe Völker, Melanie Waldenberger, Christian X. Weichenberger, Harm J. Westra, Cisca Wijmenga, Bruce H. Wolffenbuttel, Jian Yang, Patricia B. Munroe, Harold Snieder, Alan F. Wright, Igor Rudan, Laurie A. Boyer, Folkert W. Asselbergs, Dirk J. van Veldhuisen, Bruno H.Ch. Sticker, Bruce M. Psaty, Marina Ciullo, Serena Sanna, Terho Lehtimäki, James F. Wilson, Stefania Bandinelli, Alvaro Alonso, Paolo Gasparini, J. Wouter Jukema, Stefan Kääb, Vilmundur Gudnason, Stephan B. Felix, Susan R. Heckbert, Rudolf A. de Boer, Christopher Newton-Cheh, Andrew A. Hicks, John C. Chambers*, Yalda Jamshidi*, Axel Visel*, Vincent M. Christoffels*, Aaron Isaacs*, Nilesh J Samani* and Paul I.W. de Bakker*

* These authors contributed equally to this work.

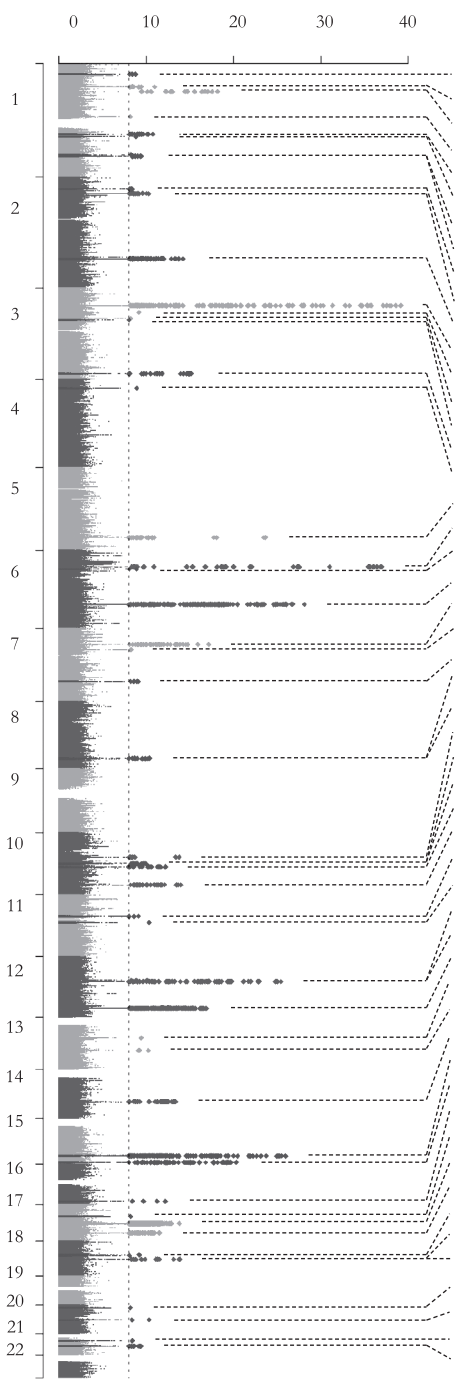
Myocardial depolarisation is a key determinant of cardiac muscle contraction and is reflected by the amplitude and duration of the QRS complex on the electrocardiogram (ECG). Increased amplitude and prolonged duration of the QRS are associated with increased electrically active myocardial mass and greater risk of heart failure and mortality. We carried out a well-powered meta-analysis of genome-wide association studies of 4 QRS traits in up to 73,518 individuals of European ancestry, followed by extensive biological and functional assessment. Here we report the identification of 52 genomic loci, of which 32 are novel, reliably associated with one or more QRS phenotypes at $P<1\times10^{-8}$, that mapped to established and putatively novel regulators of left ventricular mass. We observed enrichment in regions of open chromatin, histone modifications, and transcription factor binding in the human heart suggesting that they represent regions of the genome that are actively transcribed. We further highlighted 65 candidate genes at the identified loci that are preferentially expressed in cardiac tissue and enriched for cardiac abnormalities in *Drosophila melanogaster* and *Mus musculus*. We validated the regulatory function of a novel variant in the *SCN5A/SCN10A* locus *in vitro* and *in vivo*. Taken together, our findings provide new insights into the genetic mechanisms and biological pathways controlling electrically active myocardial mass and identify novel therapeutic targets.

The role of the heart is to provide adequate circulation of blood to meet the body's requirements of oxygen and nutrients. Cyclical depolarisation of the cardiac ventricular muscle causes contraction and results in blood flow. The QRS complex on the ECG is the most common measurement of cardiac depolarisation. Increased amplitude and duration of the QRS complex[1-3] are common alterations of myocardial depolarisation and are indicators of increased electrical activity of the left ventricular mass. Alterations in the amplitude and duration of the QRS complex are associated with clinical and preclinical cardiovascular disease such as cardiac hypertrophy and heart failure, and predict cardiovascular mortality[4,5]. Identification of specific genes influencing the QRS complex may thus lead to advances in the prevention of cardiovascular disease and death. Two previous genome-wide association studies (GWASs) have identified 22 genetic loci associated with QRS duration[6,7]. To further refine our understanding of the genetic factors influencing the QRS complex, we carried out a large scale GWAS and replication study of 4 related and clinically applied QRS traits: the Sokolow-Lyon[8,9], Cornell[9-11] and 12-lead-voltage duration products (12-leadsum)[9,11], and QRS duration[9] (Supplementary Note). Multiple ECG markers of increased left ventricular mass were examined because the sensitivity of any one of these markers alone is relatively limited and because their performance can vary with sex, ethnicity and body characteristics.[12] Our study design is summarized in Figure S1. Findings of variants convincingly associated with the QRS complex traits ($P<1\times10^{-8}$) were then examined in other ethnicities and studied for possible protein disrupting variants and co-localisation with regulatory DNA elements. Furthermore, we prioritised candidate genes by studying their cardiac gene-expression profiles and the consequences of gene knock-downs in *Drosophila melanogaster* and *Mus musculus*.

Briefly, we combined GWAS summary data from 24 studies with up to 2,766,983 autosomal SNPs. These studies together comprise 60,255 individuals of European ancestry ascertained in North America and Europe, with a maximum sample size of 54,993 for Sokolow-Lyon, 58,862 for Cornell, 48,632 for 12-leadsum, and 60,255 for QRS duration. Characteristics of participants, genotyping arrays and imputation are summarized in Supplementary Tables 1 and 2. For each SNP, evidence of association was combined across studies using an inverse-variance fixed-effect meta-analysis carried out by two independent meta-analysis groups. We performed replication testing for 35 loci showing suggestive association ($1\times10^{-8}<P<5\times10^{-7}$) in 12,838 individuals using a combination of *in silico* data and direct genotyping (Supplementary Tables 1, 2, and Supplementary Note). The threshold for genome-wide significance was set at $P<1\times10^{-8}$, allowing for a conservative Bonferroni correction for both the ~10^6 effective independent SNPs associations tested[13,14], and the 4 inter-related QRS phenotypes (Supplementary Note).

Across the genome, 52 independent loci, 32 of which are novel, reached genome-wide significance for association with one or more QRS phenotypes (Figure 1, Figure S2, and Table S3). For descriptive and downstream purposes, we defined a single 'sentinel' SNP for each locus with the lowest *P*-value against any of the four phenotypes; regional association plots for the 52 loci are shown in Figure S3. Full lists of the sentinel SNPs and the SNPs associated with any phenotype at $P<10^{-6}$ are in Supplementary Tables 3 and 4. Among the 52 known and novel loci, 32 were associated with one QRS phenotype, and 20 with two or more phenotypes (Figure S4). The total number of locus-phenotype associations at $P<10^{-8}$ was 79 (72 SNPs), of which 59 are novel (Table S5). All previously known QRS duration loci showed evidence for association ($P<10^{-6}$, Table S6). Among the 32 novel loci, 8 demonstrated genome-wide significant association with Sokolow-Lyon, 9 with Cornell, 20 with 12-leadsum, and 9 with QRS duration. Collectively, the total variance explained by the 52 sentinel SNPs for the QRS traits was between 2.7% (Sokolow-Lyon) and 5.0% (QRS-duration) (Table S7). In addition, we found evidence for 16 SNPs at 5 of the 52 loci that were not in LD with the corresponding sentinel SNP but were associated with QRS phenotypes at $P<10^{-8}$ in conditional analyses[15], suggesting independent associations at these loci (Table S8 and Supplementary Note). Among the 52 loci identified, 7 have been associated previously with PR interval (reflecting atrial activity and AV node function), 3 with QT duration (ventricular repolarisation) and 1 with heart rate (sinus node function) (Table S6), indicating at least some overlap in the genetic determinants of different cardiac measures.

Our primary GWAS and replication was based on individuals of European ancestry. To assess whether our findings are also relevant in non-European individuals, we evaluated the directional consistency of the associations in >3,600 African Americans and >4,600 Indian Asians (Figure S1, Table S9, and Supplementary note). In the African American sample, 35 of 51 available locus-phenotype associations had the same direction of effect as seen in the European sample ($P=3.19\times10^{-3}$, binomial test) and in the Indian Asian sample, 22 of 29 available locus-phenotype associations showed the same direction of effect ($P=2.91\times10^{-3}$). Thus, in concordance with observations for other traits[16], a large proportion of common variants identified in this study likely contribute to cosmopolitan genetic architecture of QRS traits.



	SENTINEL				
REGION	SNP	MAF	PHENO	P
1p36.12	rs2849028	0.26	12LS	1.69E-09
1p32.3	rs17391905	0.04	Dur	1.07E-11
1p31.3	rs2207790	0.47	Dur	6.71E-19
1p13.1	rs12039739	0.29	Dur	6.51E-09
1q22	rs2274317	0.32	12LS	1.82E-11
1q23.3	rs12036340	0.24	12LS	1.49E-09
1q32.1	rs10920184	0.38	Cor	5.01E-09
1q32.1	rs4288653	0.23	12LS	3.50E-10
2p23.3	rs6710065	0.42	Cor	4.12E-09
2p22.2	rs3770770	0.20	Dur	4.95E-11
2q31.2	rs3816849	0.46	12LS	5.77E-15
3p22.2	rs6801957	0.42	Dur	6.90E-40
3p21.1	rs4687718	0.13	Dur	6.70E-10
3p14.1	rs2242285	0.42	Dur	5.65E-09
3p14.1	rs13314892	0.23	12LS	8.93E-09
3q27.2	rs10937226	0.35	12LS	6.82E-16
4p15.31	rs1344852	0.16	Dur	1.21E-09
5q33.2	rs13185595	0.37	Cor	2.10E-24
6p21.31	rs1321311	0.27	Dur	1.03E-37
6p21.1	rs1015150	0.45	Sok	1.28E-09
6q22.31	rs11153730	0.50	Dur	7.44E-29
7p14.3	rs1419856	0.16	Dur	6.67E-18
7p12.3	rs6968945	0.44	Dur	5.14E-09
7q31.2	rs11773845	0.41	Dur	7.50E-10
8q24.13	rs4367519	0.04	Sok	4.15E-11
8q24.13	rs10105974	0.36	12LS	6.25E-11
10q21.1	rs1733724	0.26	Cor	1.75E-14
10q21.3	rs12414364	0.21	12LS	1.22E-10
10q21.3	rs10509289	0.11	12LS	9.16E-11
10q22.2	rs7099599	0.15	12LS	5.51E-13
10q25.2	rs7918405	0.26	Dur	1.05E-14
11p11.2	rs2269434	0.32	Cor	7.38E-10
11q12.2	rs174577	0.34	Dur	4.28E-11
12q13.13	rs736825	0.36	Cor	7.20E-11
12q13.3	rs2926743	0.27	12LS	3.74E-26
12q24.21	rs7132327	0.27	12LS	1.27E-17
13q14.13	rs1408224	0.31	12LS	3.60E-10
13q22.1	rs728926	0.38	Dur	5.60E-11
14q24.2	rs12880291	0.26	Dur	4.41E-14
15q25.3	rs7183401	0.44	12LS	1.10E-26
15q26.3	rs8038015	0.38	12LS	4.93E-21
16q23.3	rs6565060	0.08	12LS	6.30E-13
17q11.2	rs7211246	0.46	12LS	6.01E-09
17q21.31	rs242562	0.38	12LS	1.57E-14
17q24.2	rs9912468	0.44	Sok	3.11E-12
18q12.1	rs617759	0.33	12LS	5.63E-10
18q12.2	rs879568	0.33	Dur	8.45E-09
18q12.3	rs10853525	0.42	Dur	1.41E-14
20p12.3	rs3929778	0.20	Cor	6.42E-09
20q11.22	rs2025096	0.21	Cor	4.51E-11
21q21.1	rs7283707	0.13	12LS	3.76E-09
21q21.3	rs13047360	0.18	Dur	4.02E-10

FIG1

CANDIDATE GENES

ZNF436[n],C1orf213[n]
CDKN2C[n]
NFIA[n]
CASQ2[ng]
MEF2D[ng]
OLFML2B[n]
TNNT2[ng]
PLEKHA6[n]
DPYSL5[n]
STRN[n]
TTN[n,g]
SCN10A[ne]
TKT[n]
LRIG1[n],SLC25A26[n]
MITF[ng]
SENP2[nce]
SLIT2[n]
HAND1[ng]
CDKN1A[n,e]
TFEB[n,g]
SLC35F1[n],PLN[g],CEP85L[c]
TBX20[n,g]
TNS3[n]
CAV1[ng]
KLHL38[nc],FBXO32[g]
MTSS1[n]
DKK1[ng]
CTNNA3[ng]
CTNNA3[ng]
SEC24C[n],SYNPO2L[c]
VTI1A[n]
MYBPC3[ng],MADD[n],NR1H3[e],ACP2[c]
FADS2[n]
HOXC4[n],HOXC6[n],HOXC5[n]
NACA[nc]
TBX3[ng]
LRCH1[n]
KLF12[n]
SIPA1L1[n]
ALPK3[n]
IGF1R[n]
CDH13[ng]
NSRP1[n],EFCAB5[c]
MAPT[n]
PRKCA[n]
MAPRE2[n]
FHOD3[n,c]
SETBP1[n]
BMP2[n]
GSS[n],EDEM2[ec],MYH7B[c]
USP25[n]
ADAMTS5[n]

FIG. 2A

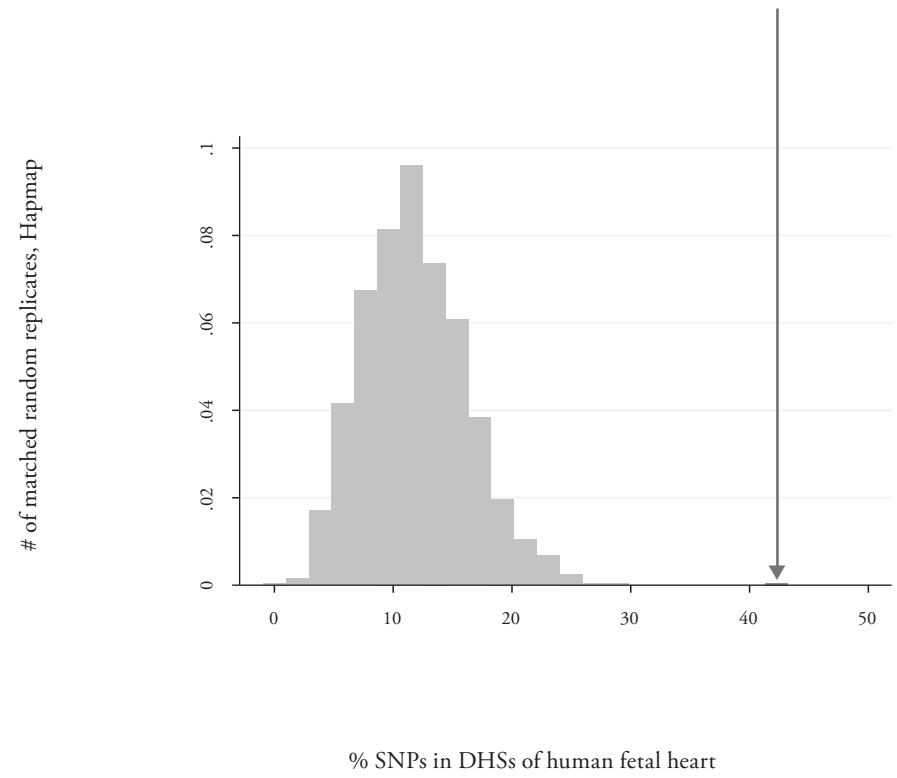
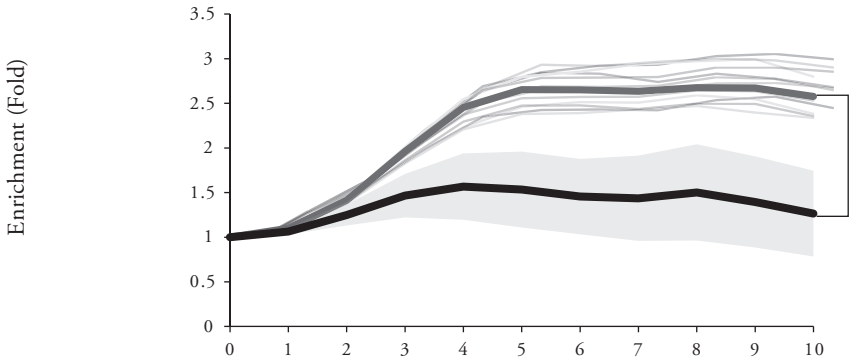


FIGURE 1
Overlay Manhattan plot showing the results for the genome-wide associations with QRS traits amongst Europeans. SNPs reaching genome-wide significance (P<1x10-8) are coloured dark grey (novel loci) or light grey (previously reported loci).

FIG.2B



Z=4.91 P=2.30E-06

FIG.2B

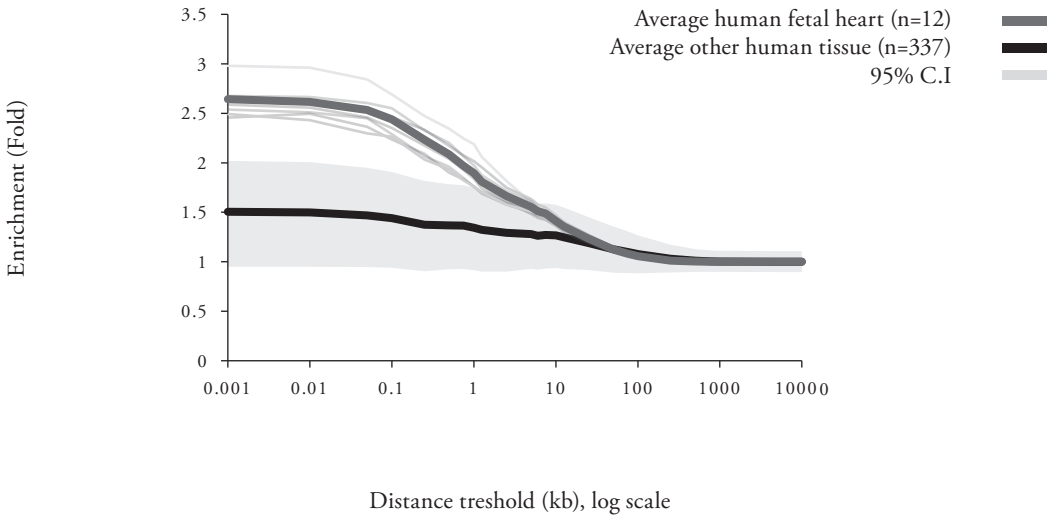


FIG.2C

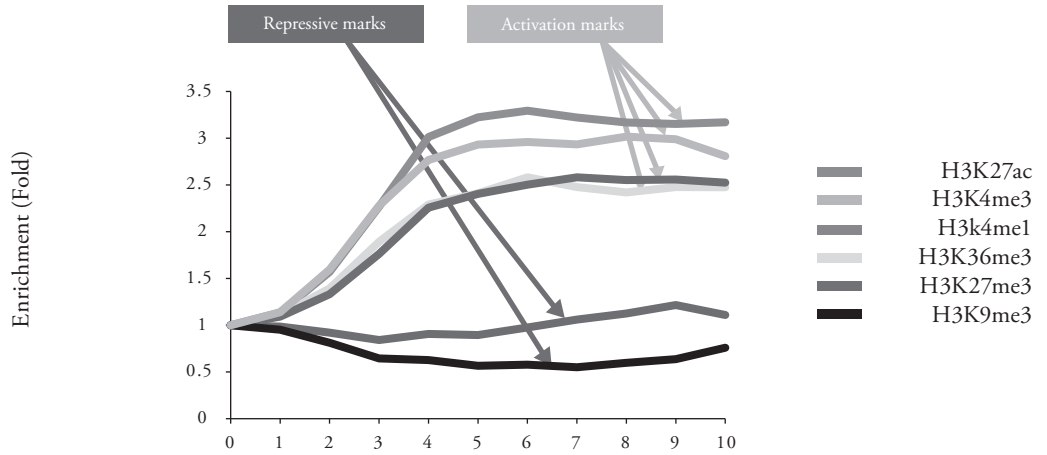


FIG.2D

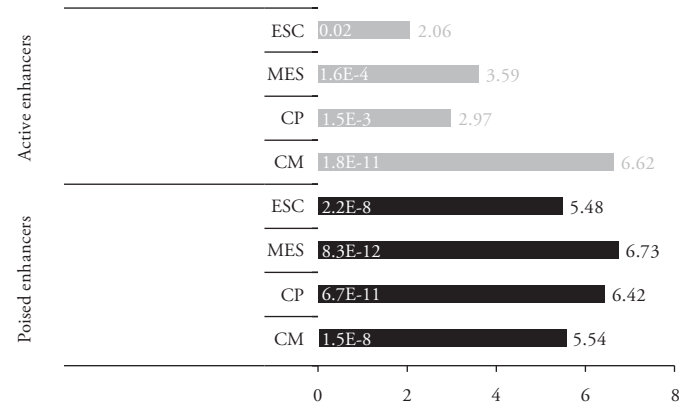


FIG.2F

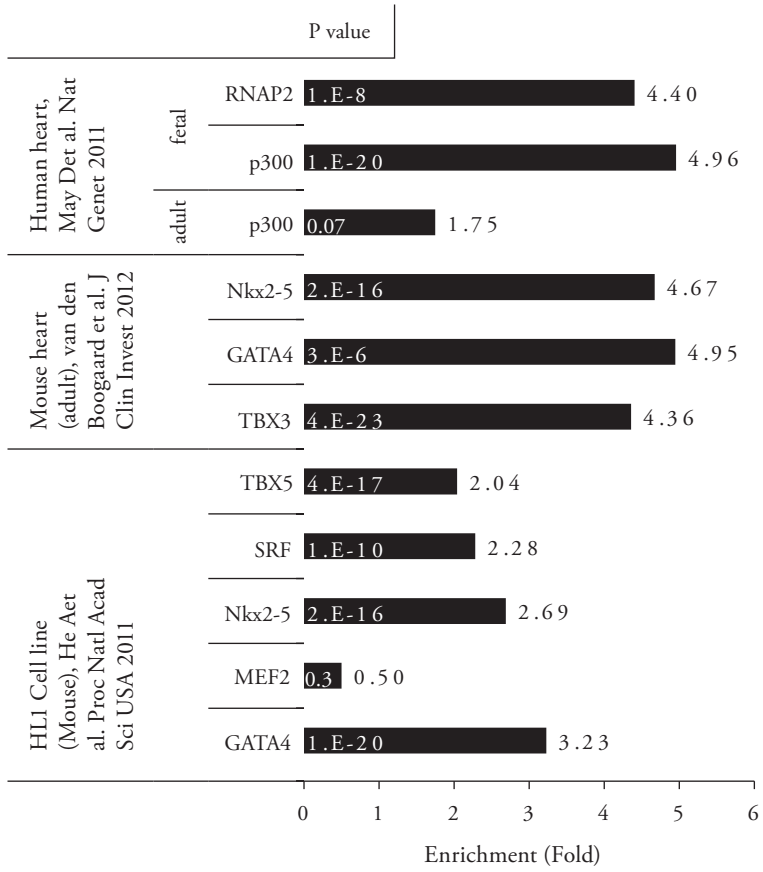


FIGURE 2.A

The 52 sentinel SNPs are significantly enriched in DHSs of the human foetal heart compared to the matched random distribution of HapMap SNPs.

B

Furthermore, we detect higher enrichment of QRS-trait SNPs in DHSs of human foetal heart (n=12) compared to all other tissues and cells (n=337) across the full range of P-values.

C

The impact of physical distance between SNPs that meet genome wide significance (P<1×10-8) on enrichment of foetal heart relative to all other tissues at DHs. The enrichment is strongest at the SNP's location and decreases after 100bp from the SNP sites.

D

SNPs associated with QRS traits are enriched for the activation histone modifications H3K27ac, H3K4me3, H3K4me1 and H3K36me3 in human left ventricle. The repressive mark H3K27me3 is not enriched while H3K9me3 is significantly reduced, suggesting that QRS-trait loci are predominantly expressed in the left ventricle.

E

Enrichment of the 52 loci for histone modifications during cardiomyocyte differentiation (mouse). Enhancers are annotated by H3K4me1 peaks at least +/- 1kb away from an annotated TSS and designated as active or poised based on the presence (active) or absence (poised) of H3K27ac.

F

SNPs (P<1×10-8) were also significantly enriched for various factors in the human heart, mouse heart and the HL-1 cell-line.

A regulatory role for QRS associations

To better capture common sequence variants at the 52 loci, we queried the 1000 Genomes Project dataset[17], and identified 41 non-synonymous SNPs in 17 genes that are in high LD (r^2>0.8) with 12 of our sentinel SNPs (Table S10). Although this is not a significant enrichment compared to the expectation under the null hypothesis (*P*=0.08, Supplementary Note), these non-synonymous sites represent an initial set of candidate variants that may have an effect on the QRS phenotypes through changes in protein structure and function. Further sequencing studies will be essential to achieve a more complete assessment of potentially causal variants. Approximately 80% of the lead signals from the combined traits were not in strong LD with known non-synonymous protein-coding variants, consistent with most of the strongest signals reported here resulting from variation in non-coding functional sequence.

To assess the importance of gene expression regulatory elements in QRS complex genetic architecture, we tested the loci identified in this study for enrichment of deoxyribonuclease I (DNase I) hypersensitive sites (DHSs), which are experimentally determined markers of regulatory activity and whose cell-type specific patterns encode early developmental fate decisions[18]. Cardiac-specific gene expression programs can be regulated by the binding of key developmental transcription factors (TFs) such as Tbx5 and Nkx2-5. When considering 349 diverse cell lines, cultured primary cells and foetal tissues [19] mapped by the ENCODE project[20] and the NIH Roadmap Epigenomics Program[21], 42 (81%) of our 52 sentinel SNPs lie in DHSs. We found that 22 (42%) of our 52 loci overlapped DHSs collected from human foetal heart tissue sampled days 96-147 after conception, representing a ~3.5-fold enrichment compared to permutation testing with 52 randomly selected SNPs matched to our sentinel SNPs ($P=7.7\times10^{-12}$, Figure 2A, Supplementary Note). This suggests a regulatory role specific to cardiac tissue. When we considered SNPs below genome-wide significance, we observed continued enrichment in DHS that diminished at the most liberal *P*-value thresholds (Figure 2B). The enrichment of genome-wide significant SNPs ($P<1\times10^{-8}$) in DHSs was strongest within the first 100 bp around the sentinel variants (Figure 2C), suggesting a specific concentration of QRS-associated variation in functional regulatory DNA.

Cardiac-specific regulatory patterns are encoded during development by sequence-specific factors such as TBX5 and NKX2-5, which recruit activating and repressive factors to impose characteristic covalent modifications of histone proteins at regulatory sites[22,23]. We investigated the enrichment of identified variants in regions of covalently modified histones in human cardiac tissue mapped by the NIH Roadmap Epigenomics Program (Supplementary Note and Table S11).[21,24] We observed a strong enrichment for histone marks associated with active enhancers, promoters and transcription (H3K4me1, H3K4me3, H3K27ac, and H3K36me3) which increased at more stringent GWAS *P*-value thresholds; by contrast no enrichment was observed for transcriptionally repressive histone marks (H3K9me3 and H3K27me3) (Figure 2D). Given the potential for genetic variants to influence the QRS complex through alterations in cardiac development, we also investigated the chromatin landscapes of mouse embryonic stem cells (ESC) differentiated in culture into mesoderm (MD), cardiac precursor (CP), and cardiomyocytes (CM) [23]. We observed that the enrichment of significant genomic regions for activating histone marks matched the degree of differentiation towards cardiomyocytes (Figure 2E). Together, these data suggest that trait-associated SNPs are enriched in regulatory elements and may alter the recruitment of sequence-specific TFs, leading to pathogenic transcriptional dysregulation[19].

To test this hypothesis, we surveyed our genome-wide significant SNPs in DHSs for perturbation of TF recognition sequences, because since these sites can point directly to binding events (Supplementary Note). Of the 22 sentinel SNPs in foetal heart DHSs, 11 are predicted to alter TF recognition sequences (Table S12). When considering all significant SNPs as well as those in high LD (r^2>0.8), 402 SNPs in the colocalising DHSs perturb transcription recognition sequences, including those of important cardiac and muscle developmental regulators like TBX, GATA-4, and MEF2. We intersected our GWAS results with ChIP-seq profiling of mouse and human cardiac tissue[25-27] and observed enrichment in enhancers marked by p300, sites bound by RNA Polymerase II (RNAP2), and the transcription factors NKX2-5, GATA-4, TBX3, TBX5, and SRF (Figure 2F). Nine of our 52 loci both have overlapping foetal heart DHSs and ChIP-seq validated TF sites. SNPs overlapping TF binding sites were 5.65 fold enriched within DHSs ($P=9.0\times10^{-10}$) but not outside DHSs (*P*=0.20).In vitro and in vivo validation of a functional cardiac enhancer

To confirm the presence of cardiac functional enhancers in our identified regions *in vivo* by direct experimentation, we examined whether enhancer candidate sequences present in these intervals are sufficient to drive reporter gene expression in the heart in transgenic mouse assays (Supplementary Note). Focusing on regions up to 200 kb away from the sentinel SNPs, we validated several candidate regions predicted by ChIP-seq as described above as reproducible *in vivo* heart enhancers. Four examples of newly identified enhancers with reproducible *in vivo* cardiac activity located in the vicinity of lead SNPs are shown in Figure 3A, additional examples of previously described enhancers near lead SNPs are provided in Figure S5 [24,26]. Recently, rs6801957 (Figure 1) in the *SCN5A/SCN10A* locus

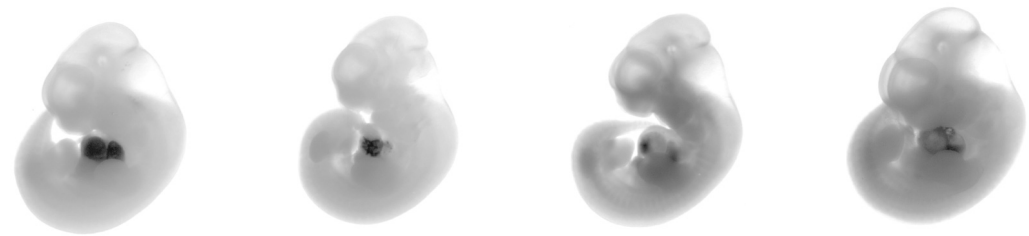
was reported to affect gene expression of another regulatory element near *SCN5A/SCN10A* through the alteration of a T-box factor binding sequence [27]. Our conditional analysis (Table S8) revealed that rs6781009 is an additional novel independent signal at this locus. This variant is located in a mouse and human heart DHS region highly occupied by activating chromatin marks and TFs (Tbx3, Tbx5 and P300) in heart tissue (Figure 3B), suggesting that this variant represents a potential additional cardiac regulatory element near *SCN5A*. Given that enhancers are thought to function by physically contacting promoters through DNA looping, we performed 4C-seq analysis oe human ventricular tissue and found an interaction between the *Scn5a* promoter and ths region around rs6781009 (Figure 3B). Evaluation of the *in vivo* activity pattern of thn regulatory element with a LacZ reporter vector showed specific expression in the interventricular septum for the major allele, which is absent for the minor allele (Figure 3C), resembling the endogenous expression pattern of both *Scn5a* and *Scn10a*[27,28]. Furthermore, functional analysis using a luciferase reporter assay in H10 cells[29] showed high constitutive activity of the human enhancer element containing the major allele for rs6781009, which was reduced when the minor allele was introduced into a large enhancer fragment or the enhancer core element (Figure 3D). Collectively, our results confirm the presence of *in vivo* heart enhancers in genome regions associated with QRS traits, and provide a starting point to identify and explore the mechanism of potentially causative variants outside protein-coding sequences.

Identification of candidate genes

Across the 52 loci, 974 annotated genes are located within 1 Mb of a sentinel SNP. Among these genes, we prioritized potential candidates using an established complementary strategy (Supplementary Note) [30,31]; we chose (i.) Genes nearest to the sentinel SNP, and any other genes within 10kb (56 genes; Figure 1); (ii.) Genes containing a non-synonymous SNP in high LD (r^2>0.8) with the sentinel SNP (11 genes; Table S10); (iii.) Genes with *cis*-eQTL associated with sentinel SNP in peripheral blood mononucleated cells (8 genes; Table S13), and (iv.) GRAIL literature analysis[32] (16 genes Table S14) with top 3 keywords retrieved from GRAIL describing the observed functional connections being 'cardiac', 'muscle', and 'heart'. In total, this strategy identified 65 candidate genes at the 52 loci (Figure 1).

Pathway analysis confirmed that the list of 65 candidate genes is strongly enriched for genes known to be involved in cardiac hypertrophy, cardiovascular disease and developmental disorder (P=1×10^{-39}; Supplementary Note)). The top 3 functional annotations for the 65 candidate genes were 'development of cardiac muscle' (7 candidate genes, P=1×10^{-9}), 'morphogenesis of cardiac muscle' (6 candidate genes, P=3×10^{-9}) and 'heart disease' (19 candi-

FIG.3A



hs2177
chr3:38594857
-38597991
SCN5A

hs2188
c h r 6:118963898
-118966469
CEP85L,PLN

hs2266
chr3:38547891-
38551036
EXOG

hs2271
chr1:26292440-
26297330

FIG.3B

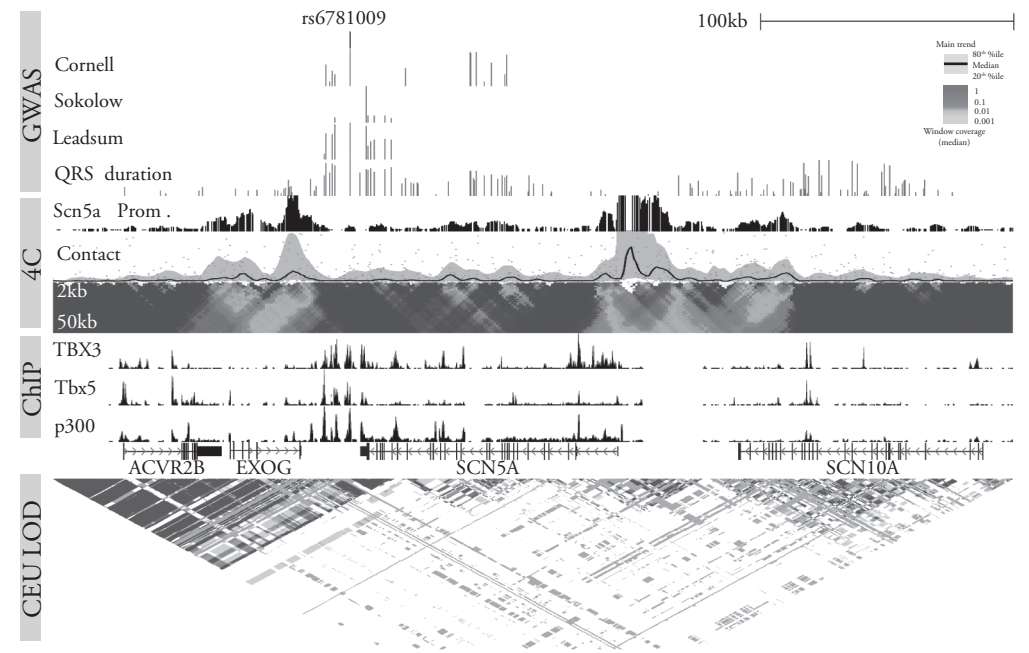


FIG.3C

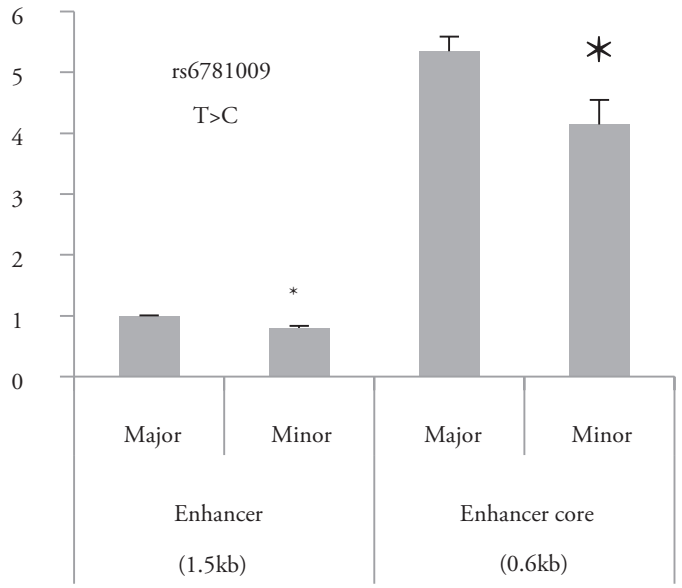


FIG.3D

FIGURE 3.A

In vivo activity of exemplar human cardiac enhancers in embryonic transgenic mice stained for LacZ enhancer reporter activity (dark blue).

B

Position of the regulatory element containing rs6781009 plotted over the GWAS signals (-log(*P*) on the *SCN5A-SCN10A* locus. The regulatory element is bound by Tbx3, Tbx5, and P300 (lower black traces) in mouse, and the contact profile of the Scn5a promoter obtained by 4C-seq in human ventricular tissue revealed an interaction between this regulatory element and the Scn5a promoter (upper black trace and contact profile). Normalized contact intensities (gray dots) and their running median trends (black line) are depicted for the Scn5a promoter viewpoint. Medians are computed for 4 kb windows and the grey band displays the 20–80% percentiles for these windows. Below the profile, statistical enrichment across differently scaled window sizes (from 2 kb (top row) to 50 kb (bottom)) is depicted of the observed number of sequenced ligation products over the expected total coverage of captured products, with the latter being estimated based on a probabilistic background model. Local changes in colour codes indicate regions statistically enriched for captured sequences.

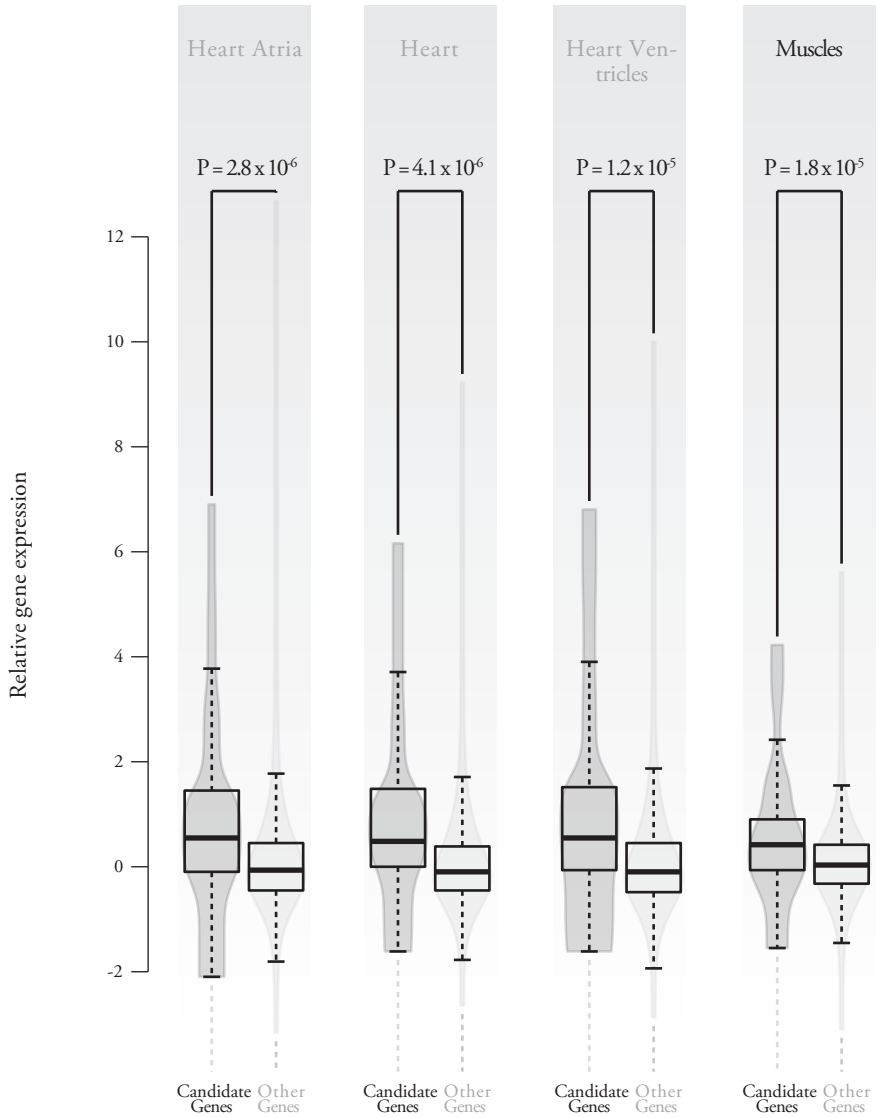
C

Luciferase assay performed in H10 cells showing a high constitutive activity for the enhancer core element (0.6kb) containing the major allele for rs6781009, which is reduced for the minor allele in both a large enhancer construct (1.5kb), as well as in the core enhancer element (0.6kb) *P<0.01

D

Dorsal views of hearts containing the human regulatory element with the major vs minor allele for rs6781009 in a LacZ reporter vector, showing specific expression of the enhancer in the interventricular septum (ivs, black bars) for the major allele, which is absent for the minor allele *P<0,05. ra, right atrium; la, left atrium; rv, right ventricle; lv, left ventricle.

FIG.4A


Heart Atria
P = 2.8 x 10-6
Heart
P = 4.1 x 10-6
Heart Ventricles
P = 1.2 x 10-5
Muscles
P = 1.8 x 10-5
Relative gene expression
12
10
8
6
4
2
0
-2
Candidate Genes
Other Genes

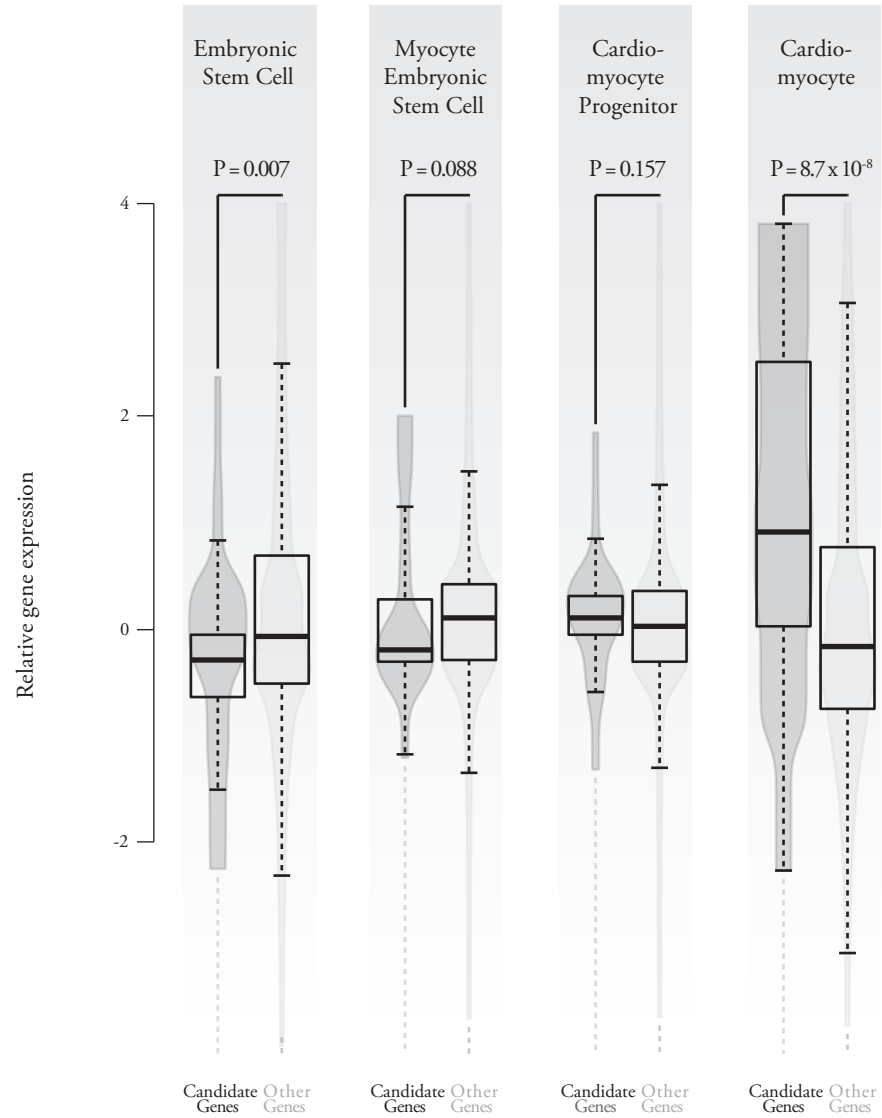

Embryonic Stem Cell
P = 0.007
Myocyte Embryonic Stem Cell
P = 0.088
Cardio-myocyte Progenitor
P = 0.157
Cardio-myocyte
P = 8.7 x 10-8
Relative gene expression
4
2
0
-2
Candidate Genes
Other Genes

FIG.4B

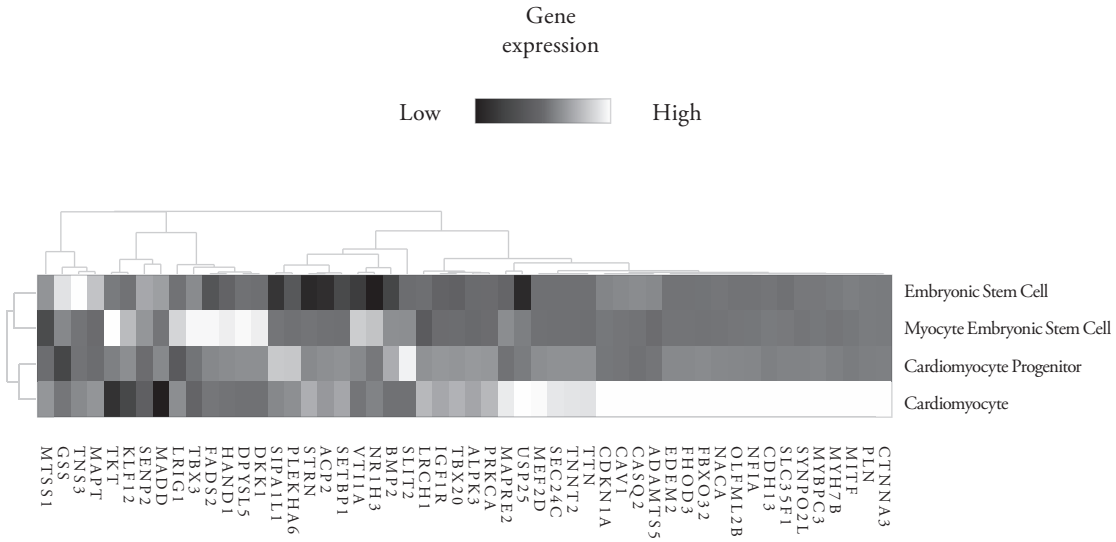


FIGURE 4.A

Within heart and muscle tissue microarray-based data the candidate genes are significantly more highly expressed as compared to non-candidate genes.

B

The 54 candidate genes are highly expressed in RNA-seq data of cardiomyocytes, compared to non-candidate genes.

C

Unsupervised hierarchical clustering of RNA-seq based expression data of 54 candidate genes in 4 different cardiomyocyte (precursors) reveals that most of the genes are abundantly expressed in cardiomyocytes.

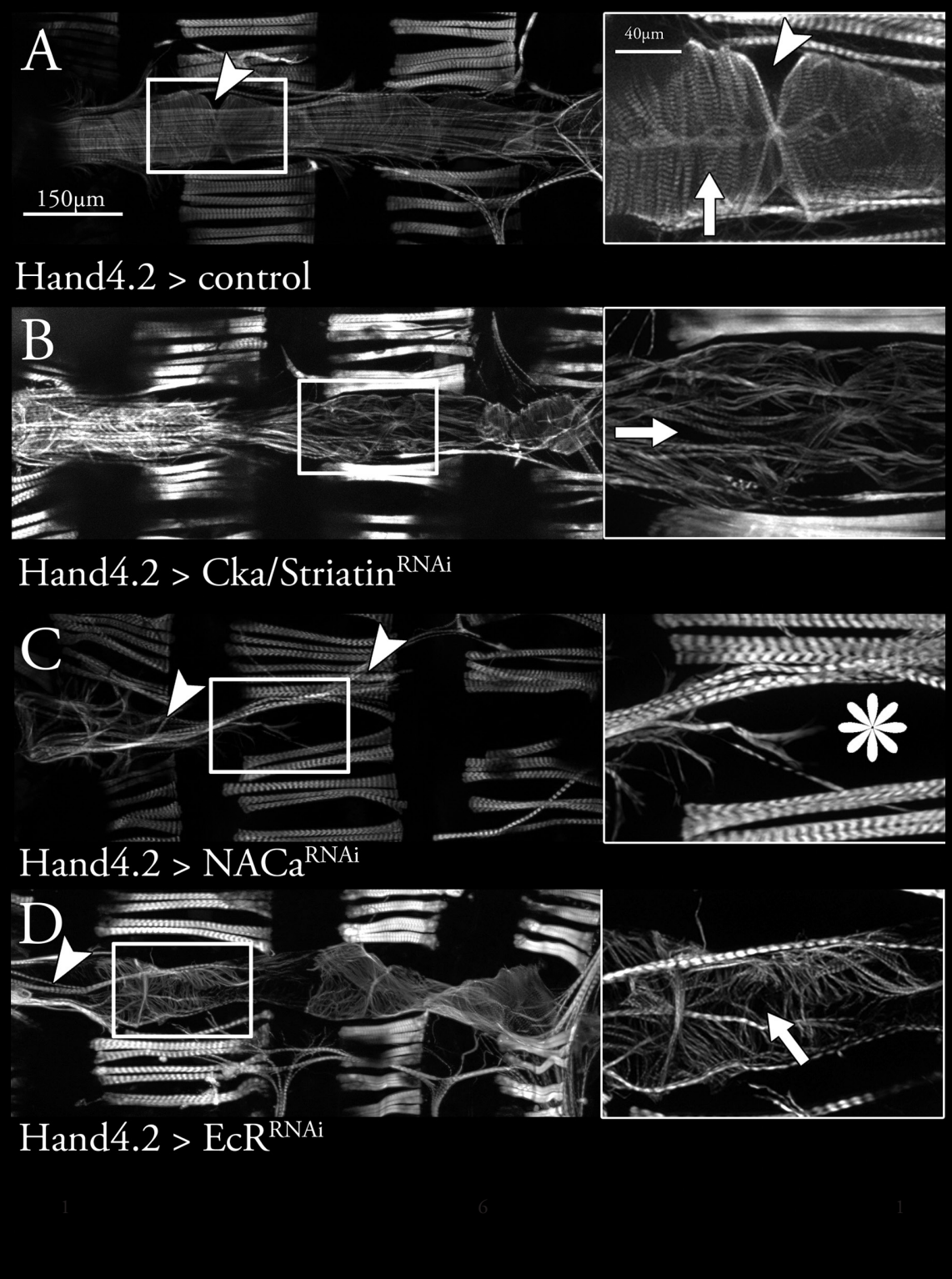
date genes, P=8×10^{-9}) (Supplementary Tables 15 and 16). Current knowledge on gene function for all 65 candidates is summarized in Table S17. A systematic search in the Online Mendelian Inheritance in Man (OMIM) catalogue revealed that several candidate genes are known to cause familial cardiomyopathies (*TNNT2*, *TTN*, *PLN*, *MYBPC3*) or cardiac arrhythmias (*CASQ2*) in humans (Table S17). We also identified genes that are associated with atrial septal defects (*TBX20*) and more complex syndromes involving cardiac abnormalities such as the Schinzel-Giedion midface retraction syndrome (*SETBP1*)[33] and the Ulnar-mammary syndrome (*TBX3*)[34].

Insights from gene expression profiling and model organisms

We explored gene expression profiles of our candidate genes in data derived from 37,427 Affymetrix U133 Plus 2.0 arrays across 40 annotated tissues (Supplementary Note). Of our 65 candidate genes we could reliably assign a probe for 63 genes; these transcripts were on average expressed at higher levels in cardiac-derived samples compared to other transcripts in the same sample (P=4.1×10^{-6} for heart tissue; Wilcoxon test; Figure 4A). Further, expression of these transcripts was expressed at higher levels in cardiac-derived samples compared to other tissues (P=0.006 after Bonferroni correction; Figure S6). To further investigate the potential role of these candidate genes in cardiac development, we assessed temporal gene expression patterns during *in vitro* differentiation of mouse embryonic stem cells (ESC) via mesoderm (MD) and cardiac precursor (CP) cells to cardiomyocytes (CM). Eight percent of genes are mainly expressed during the ESC stage, 19% during MD stage, 8% in the CP stage and 62% in the cardiomyocyte stage (Figure 4B). Compared to other genes, the candidate genes are high expressed in cardiomyocytes (P=4.7×10^{-8}; Wilcoxon test; Figure 4C). These results suggest that our candidate gene set is enriched for genes relevant to cardiac biology, and include a number of genes differentially expressed in cardiac tissue and increasingly expressed during cardiac development.

Next, we analysed data from model organisms to explore the function of the selected candidate genes (Supplementary Note). From cardiac tissue-specific RNAi knockdown data collected in *D. melanogaster*[35], we found that the 65 candidate genes were 2.3-fold enriched for stress-induced premature cardiac death (9 genes, P=5.4×10^{-3}; Figure S7A). Four of these genes had been studied previously in *Drosophila* and shown to have cardiac abnormalities in *Drosophila* (*Mhc*/MYH7B[36], *Slit*/SLIT2[37], *EcR*/NR1H[38], *Hand*/HAND1[39]). Performing heart-specific RNAi knockdown with the cardiac Hand4.2-Gal4 driver line (Supplementary Note), we re-tested *EcR*, which has multiple homologous genes in mammals, and *Hand* as well as the remaining four genes where

FIG.5



1 6 1

Figure 5.
Cardiac defects upon heart-specific RNAi knockdown in Drosophila . (a) Wild-type dorsal heart tube stained with the F-actin stain phalloidin. Magnified region (right) is highlighted. Arrowheads point to ostia (inflow tracks), arrow shows the circumferential orientation of myofibrils. (b) Cka/Striatin RNAi induces myofibrillar disarrangement. Myofibrils are oriented in a disorganized, mainly anterior-posterior orientation with gaps in between (arrow). (c) Knockdown of NACa/NACA causes severe cardiac tissue disintegration. Adult cardiomyocyte tissue may be complete absent (asterisk), while some heart-associated longitudinal muscles are still present (arrowheads). At larval stages the heart is much less affected, suggesting a maturation or remodeling defect. (d) Knockdown of EcR/NR1H blocks cardiac remodelling and causes myofibrillar disarray (arrow). Ventral longitudinal muscles are also abnormal (arrowhead).

the cardiac phenotype was not known (we did not test Titin, since it is a well-known cardiomyopathy gene[40]). Adult hearts of *EcR*/NR1H, *NACa*/NACA, *Hand*/HAND1 and Cka/STRN RNAi showed severe cardiac defects (Figure 5). Knockdown of *Hand*/HAND1 and *Cka*/STRN both had a reduced cardiac heart rate (Figure S8). While *Hand*/HAND1 knockdown hearts appeared structurally normal, we observed severely disorganized and misoriented myofibrillar arrangements within the cardiomyocytes in *Cka*/STRN RNAi hearts (Figure 5), which caused a reduction in diastolic diameters and fractional shortening (Suppl. Figure S7C). NACa/NACA mutants had the most severe phenotype with a complete loss of cardiac tissue beginning at eclosion, while the hearts of NACa mutant larvae were still intact, indicating a critical role for NACA during cardiomyocyte remodelling. RNAi-mediated knockdown of *CG4743*/SLC25A26 and *Fhos*/FHOD3 did not reveal cardiac phenotypes.

In addition, from the Mouse Genome Informatics database, knockout models were annotated for 44 of the 65 orthologous genes, of which 18 (41%) revealed a cardiac phenotype (Table S17). This represents a 5.4-fold enrichment compared to randomly matched sets of 65 genes ($P=4.8\times10^{-16}$; Figure S7B). This further demonstrates the relevance of these genes for cardiac function.

For some loci, there are additional candidate genes. A notable example is the 11p11.2 locus, which harbours multiple candidate genes (Figure 1), including *MYBPC3*, *ACP2*, *MADD*, and *NR1H3*. *MYBPC3* deficiency is well established to cause hypertrophic and dilated cardiomyopathies in both human and mouse models and provides a plausible candidate gene (Table S17). In addition to *MYBPC3*, eQTL and histone modification data also suggests a potential role for *NR1H3* (Figure S9). Decreased expression of *NR1H3* was associated with higher QRS voltages. However, *NR1H3* deficient mice do not spontaneously develop a cardiac hypertrophic phenotype (MGI: 1352462). To study the potential cardiac effects of *NR1H3* overexpression, we created a transgenic mouse with cardio-specific overexpression of *NR1H3* under the control of the *Myh6* promoter. While the wild-type mouse was sensitive to perturbations (such as transverse aortic constriction and angiotensin 2 infusion) that provoke cardiac hypertrophy[41], these effects were absent in the transgenic mouse. This observation is in line with protective effects due to treatment with T0901317, a synthetic *NR1H3* agonist, in mice challenged with aortic constriction[42]. These data highlight the importance of systematic approaches to identify causal genes beyond identification of a first recognisable candidate.

CONCLUSIONS

We report a meta-analysis of GWAS in 73,518 individuals for 4 quantitative QRS phenotypes and identify 52 independent genetic loci influencing these traits with 79 locus-phenotype associations; the majority of these discoveries are novel. Our analyses prioritized a set of candidate genes that potentially impact cardiac function, which we expect to facilitate in-depth studies towards identifying definitive mechanisms. Our loci co-localised with open chromatin, histone modification, and TF binding sites specifically in cardiac tissue, and provide examples of *in vivo* functional enhancers within the identified loci. We also provide direct evidence that rs6781009, located within a predicted enhancer, interacts with the promoter of *SCN5A* to modify expression levels. In parallel, we have prioritized a set of 65 genes based on differential gene expression in the human heart, making them promising candidates for mediating effects on QRS phenotypes. Further functional support is obtained from genetic overlap with inherited Mendelian disorders and from model organism studies. We anticipate that in-depth functional studies will now detail mechanisms affecting cardiac function and advance understanding of cardiac hypertrophy and heart failure.

METHODS

A summary of the methods can be found in Supplementary Information and includes detailed information on: study populations; genotyping methods and quality control; electrocardiographic measurements; genome-wide association and meta-analysis methods; gene prioritisation strategies; experimental data sets and analytical methods; *in vivo* and *in vitro* experiments.

LITERATURE

1. Levy, D. et al. Determinants of sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy. *Circulation* 81, 815-20 (1990).

2. Devereux, R.B., Koren, M.J., de Simone, G., Okin, P.M. & Kligfield, P. Methods for detection of left ventricular hypertrophy: application to hypertensive heart disease. *Eur Heart J* 14 Suppl D, 8-15 (1993).

3. Okin, P.M. et al. Time-voltage QRS area of the 12-lead electrocardiogram: detection of left ventricular hypertrophy. *Hypertension* 31, 937-42 (1998).

4. Kannel, W.B., Gordon, T. & Offutt, D. Left ventricular hypertrophy by electrocardiogram. Prevalence, incidence, and mortality in the Framingham study. *Ann Intern Med* 71, 89-105 (1969).

5. Verdecchia, P. et al. Prognostic value of a new electrocardiographic method for diagnosis of left ventricular hypertrophy in essential hypertension. *J Am Coll Cardiol* 31, 383-90 (1998).

6. Sotoodehnia, N. et al. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat Genet* 42, 1068-76 (2010).

7. Holm, H. et al. Several common variants modulate heart rate, PR interval and QRS duration. *Nat Genet* 42, 117-22 (2010).

8. Sokolow, M. & Lyon, T.P. The ventricular complex in left ventricular hypertrophy as obtained by unipolar precordial and limb leads. *Am Heart J* 37, 161-86 (1949).

9. Carlsson, M.B. et al. Left ventricular mass by 12-lead electrocardiogram in healthy subjects: comparison to cardiac magnetic resonance imaging. *J Electrocardiol* 39, 67-72 (2006).

10. Casale, P.N. et al. Electrocardiographic detection of left ventricular hypertrophy: development and prospective validation of improved criteria. *J Am Coll Cardiol* 6, 572-80 (1985).

11. Molloy, T.J., Okin, P.M., Devereux, R.B. & Kligfield, P. Electrocardiographic detection of left ventricular hypertrophy by the simple QRS voltage-duration product. *J Am Coll Cardiol* 20, 1180-6 (1992).

12. Hancock, E.W. et al. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part V: electrocardiogram changes associated with cardiac chamber hypertrophy: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. *Circulation* 119, e251-61 (2009).

13. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-78 (2007).

14. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 32, 381-5 (2008).

15. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 44, 369-75, S1-3 (2012).

16. Teslovich, T.M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466, 707-13 (2010).

17. Abecasis, G.R. et al. A map of human genome variation from population-scale sequencing. *Nature* 467, 1061-73 (2010).

18. Stergachis, A.B. et al. Epigenetic memory of developmental fate and time encoded in human regulatory DNA landscapes. *Cell* 154, 888-903 (2013).

19. Maurano, M.T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337, 1190-5 (2012).

20. Thurman, R.E. et al. The accessible chromatin landscape of the human genome. *Nature* 489, 75-82 (2012).

21. Bernstein, B.E. et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 28, 1045-8 (2010).

22. Miller, S.A., Mohn, S.E. & Weinmann, A.S. Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. *Mol Cell* 40, 594-605 (2010).

23. Wamstad, J.A. et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* 151, 206-20 (2012).

24. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74 (2012).

25. He, A., Kong, S.W., Ma, Q. & Pu, W.T. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci U S A* 108, 5632-7 (2011).

26. May, D. et al. Large-scale discovery of enhancers from human heart tissue. *Nat Genet* 44, 89-93 (2012).

27. van den Boogaard, M. et al. Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *J Clin Invest* 122, 2519-30 (2012).

28. Remme, C.A. et al. The cardiac sodium channel displays differential distribution in the conduction system and transmural heterogeneity in the murine ventricular myocardium. *Basic Res Cardiol* 104, 511-22 (2009).

29. Jahn, L. et al. Conditional differentiation of heart- and smooth muscle-derived cells transformed by a temperature-sensitive mutant of SV40 T antigen. *J Cell Sci* 109 (Pt 2), 397-407 (1996).

30. Gieger, C. et al. New gene functions in megakaryopoiesis and platelet formation.

Nature 480, 201-8 (2011).

31. van der Harst, P. et al. Seventy-five genetic loci influencing the human red blood cell. *Nature* 492, 369-75 (2012).

32. Raychaudhuri, S. et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* 5, e1000534 (2009).

33. Hoischen, A. et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 42, 483-5 (2010).

34. Linden, H., Williams, R., King, J., Blair, E. & Kini, U. Ulnar Mammary syndrome and TBX3: expanding the phenotype. *Am J Med Genet A* 149A, 2809-12 (2009).

35. Neely, G.G. et al. A global in vivo Drosophila RNAi screen identifies NOT3 as a conserved regulator of heart function. *Cell* 141, 142-53 (2010).

36. Melkani, G.C., Bodmer, R., Ocorr, K. & Bernstein, S.I. The UNC-45 chaperone is critical for establishing myosin-based myofibrillar organization and cardiac contractility in the Drosophila heart model. *PLoS One* 6, e22579 (2011).

37. Qian, L., Liu, J. & Bodmer, R. Slit and Robo control cardiac cell polarity and morphogenesis. *Curr Biol* 15, 2271-8 (2005).

38. Monier, B., Astier, M., Semeriva, M. & Perrin, L. Steroid-dependent modification of Hox function drives myocyte reprogramming in the Drosophila heart. *Development* 132, 5283-93 (2005).

39. Han, Z., Yi, P., Li, X. & Olson, E.N. Hand, an evolutionarily conserved bHLH transcription factor required for Drosophila cardiogenesis and hematopoiesis. *Development* 133, 1175-82 (2006).

40. Herman, D.S. et al. Truncations of titin causing dilated cardiomyopathy. *N Engl J Med* 366, 619-28 (2012).

41. Cannon, M.V. et al. Cardiac LXRα overexpression protects against pathological hypertrophy and dysfunction by enhancing glucose uptake and utilization. (2014).

42. Kuipers, I. et al. Activation of liver X receptors with T0901317 attenuates cardiac hypertrophy in vivo. *Eur J Heart Fail* 12, 1042-50 (2010).